Discriminant analysis for the prediction and classification of tick-borne infections in some dairy cattle herds at Dakahlia Governorate, Egypt

Eman A. Abo El Fadl¹,§, Maged El-Ashker²,§, Keisuke Suganuma³, ⁴ and Mitsunori Kayano³, *

¹Department of animal husbandry and development of animal wealth (biostatistics), Faculty of Veterinary Medicine, Mansoura University, Mansoura 35516, Egypt.
²Department of Internal Medicine and Infectious Diseases, Faculty of Veterinary Medicine, Mansoura University, Mansoura 35516, Egypt.
³Research Center for Global Agromedicine, Obihiro University of Agriculture and Veterinary Medicine, Inada-cho, Obihiro, Hokkaido, 080–8555, Japan
⁴National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Inada-cho, Obihiro, Hokkaido, 080–8555, Japan

Received for publication, December 26, 2016; accepted, April 25, 2017

Abstract
This study was undertaken to use the variable loadings in linear discriminant analysis (LDA) to determine the most important predictors for the discrimination of tick-borne diseases (TBDs), particularly babesiosis and anaplasmosis and predict the group membership from the predictors. In total, 163 cattle, from different localities at Dakahlia Governorate, Egypt, were investigated in 2012 and 2013 for the presence of TBDs. All cattle were clinically examined and a clinical index score was determined for each cow. Blood samples were also collected from each animal for adopting microscopy and diagnostic laboratory methods. Out of the examined cattle, 83 animals were acutely-ill (Babesia bovis and Anaplasma marginale were identified in 11 and 10 animals, respectively), while 80 cows were apparently healthy but having previous attacks of blood parasites (23 animals harbored anaplasma marginale (asymptomatic carriers)). The remained 119 animals were negative to TBDs. Fourteen animals were not survived and 149 cases were survived. As the result of the first LDA to discriminate babesiosis, anaplasmosis and negative to TBDs, 89.0% of animals were correctly classified; 78.8% (26/33) for anaplasma, 100% (11/11) for babesia infections, 90.8% (108/119) for negative to TBDs, respectively. The important predictors for the discrimination were oculonasal discharge, bloody feces, hemoglobinuria, bloody feces and respiratory rate. On the other hand, the second LDA discrimination showed high classification accuracy of 87.1% for the discrimination of survivors and non-survivors; 89.9% (134/149) for survivors and 57.1% (8/14) for non-survivors, while the important predictors included oculonasal discharge, recumbent posture and nervous sign.

Key Words: Classification; discriminant analysis; pathogens; predictors; tick-borne

§Equally contributed
*Corresponding author: Mitsunori Kayano, Research Center for Global Agromedicine, Obihiro University of Agriculture and Veterinary Medicine, Inada-cho, Obihiro, Hokkaido, 080–8555, Japan
Phone: +81–155–49–5521. Fax: +81–155–49–5593. E-mail: kayano@obihiro.ac.jp
doi: 10.14943/jjvr.65.3.127
Introduction

Tick-borne diseases (TBDs), caused by Babesia bovis, B. bigemina and Anaplasma marginale, are crucial infections of cattle in tropical and sub-tropical regions in Egypt\(^1,2,4\). These infections cause constraints to the livestock improvement programs in Egypt with resulting a decrease of animal productivity and extreme economic losses. Animals suffering from acute attacks of babesiosis or anaplasmosis can pertain similar clinical presentation. Although the associating clinical symptoms are not pathognomonic, animals with chronic infections can be asymptomatic carriers\(^1,9,11\). Hence, accurate diagnosis as well as the prediction of the clinical outcome of these infections is required for appropriate and immediate treatment after the infection.

Discriminant analysis (DA) is a statistical technique used for differentiating groups (categorical dependent variable) by using the various independent and typically quantitative variables (parameters). Linear discriminant analysis (LDA) is mainly used in statistics (1) to find a linear combination of variables that classifies objects or events into two or more classes and (2) to identify important variables for the classification\(^8,14,15\). LDA allows us to predict a group membership based on a linear combination of the independent variables, hence careful examination of the prediction model that results from the procedure can give an insight into the relationship between group membership and the variables used to predict group membership\(^16\).

In previous reports, it has shown that DA could be used for discriminating the type of pathogens causing infection and to determine the best predictors for classification. It had been used to analyze data of eleven variables involving (red blood cells, packed cell volume, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin content, hemoglobin concentration, white blood cells, neutrophils, leukocytes, platelets, rectal temperature), to identify which parameters out of the eleven are the most reliable parameters to be considered as markers of the ovine anaplasmosis\(^3\).

On the other side, the forward stepwise discriminant function analysis (DFA) have been used for classification of the groups of Atlantic cod (Gadus morhua) by using several parasites as “parasite tags”\(^6\). In another report, DA on the morphometric data of opisthaptor to determine which parameters more accurately distinguish between Diplectanum cazauxi and Diplectanum bauchotae has been reported and indicated that the best discriminant parameters are the copulatory apparatus and the central bar which determined the absence of misclassification in the model\(^7\).

Up to now, there have been limited data about the value of using discriminant analysis in cattle with TBDs. In line with these considerations, the present study was delineated to check the significance of different determinants to discriminate the type of infection and to predict the clinical outcomes of the affected subjects.

Materials and Methods

Animal population and data collection: The present study included 163 cattle (130 from 6 dairy farms; 33 small-holders) that were located at Dakahlia Governorate, Egypt. The original study included 165 animals, however two animals were not follow the study design and thus were excluded. The animals were investigated for the presence of one or more TBDs during summers 2012 and 2013. The ages ranged between 1–5 years in dairy farms and between 6 months and 2-years for the animals of small-holders. The 83 animals, of four dairy farms (\(n = 50\)) as well as those of small-holders (\(n = 33\)), had recent attacks of babesiosis or anaplasmosis with variable degrees of tick infestation as well as sporadic cases of sudden deaths in the respective herds (acute cases). These cattle typically had pyrexia, pale to icteric visible mucosa, respiratory
embarrassment, and oculonasal discharge. Some rare cases showed enlarged superficial lymph nodes and the others had red urine. In the remained two farms (n = 80), no clinical signs appeared upon examination but a previous outbreak was the main complaint of the enclosed owners (chronic cases). Out of the 83/163 acute cases, 11 and 10 animals were infected with babesiosis and anaplasmosis respectively, while 23 chronic animals out of 80/163 were positive for anaplasmosis. Fourteen infected animals were not survived, and 149 cases were survived.

All cattle were clinically examined and a clinical index score was determined for each cow. The scoring system was based on clinical and laboratory variables including rectal temperature, heart rate, respiratory rate, coughing, dyspnea, mucous membrane color, superficially located lymph nodes, oculonasal discharge, color of urine, appetite, blood in feces, nervous signs, posture, frothy salivation and packed cell volume. Microscopic examination of Giemsa stained blood smears was performed for all investigated cattle.

**Research hypothesis:**

$H_0$: Any linear combination of the predictors has no discriminating ability to tick-borne disease.

$H_A$: There exists at least one linear combination of the predictors has discriminating ability to tick-borne disease.

**Statistical analysis:**

**Discriminant analysis (DA)**

DA is a statistical multivariate technique used in many different fields. LDA involves a discriminant variety and represents a linear combination of two or more predictors that discriminate between the objects in the groups defined a priori. The assumptions of LDA can be summarized as: Homogeneous within the group variances, multivariate normality within group, and linearity among all pairs of variables, no multi-collinearity was detected.

Data were analyzed by using SPSS version 16, USA. Two LDA models were used. The first model was carried out to check the significance of different determinants to discriminate the type of TBDs ($Anaplasma$, $Babesia$) as dependent variables by using clinical signs as independent variables. The type of infection (dependent variable) was coded before starting the analysis as following ($Anaplasma = 1$, $Babesia = 2$ and negative to TBDs = 3). All tested clinical findings including rectal temperature, heart rate, respiratory rate, cough, dyspnea, mucous membranes color, enlarged lymph nodes, oculonasal discharge, hemoglobinuria, appetite, blood feces, nervous signs, posture, frothy salivation were then entered in DFA model using stepwise method and from all of these variables only respiratory rate, hemoglobinuria, bloody feces, oculonasal discharge, mucous membrane and size of lymph node were selected to be in the equation.

The second LDA was carried out to check the significance of different determinants to predict the clinical outcomes of the affected subjects (survivors vs. non survivors) using clinical outcomes as dependent variables and the clinical findings as independent variables. The dependent variable the affected subjects was coded as following (survivors = 1 and non survivors = 0). All examined clinical signs including rectal temperature, heart rate, respiratory rate, cough, dyspnea, mucous membranes color, enlarged lymph nodes, oculonasal discharge, hemoglobinuria, appetite, blood feces, nervous signs, posture, frothy salivation were then entered in DFA model using stepwise method and from all of these variables only oculonasal discharge, posture and nervous signs were selected to be in the equation. As described in the first LDA model. For both models, Wilk’ lambda, group centroids and classification accuracy were calculated.

The discriminant statistical model used for this analysis was:

$$DF = V_1X_1 + V_2X_2 + V_3X_3 + ... + V_pX_p,$$

Where $DF =$ discriminate function (score)
of grouping variables, \( V = \) the standardized discriminant coefficient or loadings for the clinical signs (predictors), \( X = \) respondent’s score for the clinical signs, \( I = \) the number of predictor variables. The discriminant function coefficients \( V \) or standardized form beta indicate the partial contribution of each clinical signs to the discriminate function controlling for all other variables in the equation. They can be used to assess unique contribution to the discriminate function and therefore provide information on the relative importance of each variable. They also calculate the discriminant score for a given case. The score is calculated in the same manner as a predicted value from a linear regression. The magnitudes of these coefficients indicate how strongly the discriminating variables affect the score. They allow comparing variables measured on different scales. Coefficients with large absolute values correspond to variables with greater discriminating ability.

**Results**

In the present study, the initial presumptive diagnosis of both bovine babesiosis and anaplasmosis was achieved on the basis of case history, clinical symptoms, and microscopy, while confirmation was done using PCR assays and analysis of gene sequences. Six clinical variables including respiratory rate, hemoglobinuria, bloody feces, oculonasal discharge, mucous membrane, and size of lymph node were selected by DFA in first LDA, while three clinical finding including recumbent posture, nervous signs and oculonasal discharge were likely used for discrimination process in second LDA model (Table 1 and 2).

It was found that 89.0% of animals were correctly classified by using first LDA model. The average rate of correct classification was 78.8% (26/33) for *Anaplasma*; 100% (11/11) for *Babesia* infections (Table 3). On the other hand, the average rate of correct classification for survivors was 89.9% (134/149), whereas it was 57.1% (8/14) for non-survived subjects with 87.1% percent of correctly classified samples. The low accuracy of discrimination for non-survivors (57.1%) is due to not all dead animal showed abnormal posture or nervous signs so, these cases led to decrease the discrimination accuracy for non-survivors animals. The values of the statistical parameters F and Wilk’s lambda for the discriminant functions showed a reasonable statistical significance. For the first LDA model, the first discriminant functions were identified to discriminate the type of infection. The function explained 94.2% of the variance with 0.058 Wilks’ Lambda while the second one showed 62.4% of the variance with 0.376 Wilks’ Lambda (Table 1). The first discriminant functions identified oculonasal discharge, bloody feces and hemoglobinuria (1.74, \(-0.894\) and 0.576) as the important predictors for the first function to discriminate anaplasmosis from negative to TBDs, while bloody feces and respiratory rate were the best predictors for the second one for identifying the type of infection as these variables had large coefficients (\(-0.850, 0.728\)), respectively. Whereas score for other variables was less successful for prediction as these variables had small coefficients so it considered low predictors.

**The first LDA model:**

\[
\begin{align*}
DF1 &= -0.282^{*}\text{respiratory rate} \\
&\quad - 0.495^{*}\text{lymph node} \\
&\quad - 0.434^{*}\text{mucous membrane} \\
&\quad + 1.735^{*}\text{oculonasal} \\
&\quad + 0.576^{*}\text{Hemoglobinuria} \\
&\quad - 0.894^{*}\text{bloody faces},
\end{align*}
\]

\[
\begin{align*}
DF2 &= 0.728^{*}\text{respiratory rate} \\
&\quad - 0.373^{*}\text{lymph node} \\
&\quad + 0.620^{*}\text{mucous membrane} \\
&\quad - 0.252^{*}\text{oculonasal} \\
&\quad + 0.295^{*}\text{Hemoglobinuria} \\
&\quad - 0.850^{*}\text{bloody faces}.
\end{align*}
\]

Using this function, the discriminant scores for all cases were calculated. Cases with score near
to the group centroid are predicted as belonging to that group, so any new cases being added can be classified (Table 2). For the second LDA model, one function was identified which explained 34.5% of the variance with 0.655 Wilks’ Lambda. This function identified oculonasal discharge, recumbent posture and nervous sign (0.558, 0.510 and 0.464) as important or strong predictor to discriminate survivors from non-survivors as these variables had large coefficients; whereas score of other variables were less successful for prediction as these variables had small coefficients, thus they considered low predictors. From this function we calculated the discriminant scores for all cases and compared with the group centroid (Table 2) as previously mentioned.

**Second LDA model:**

\[
DF2 = 0.558 \times \text{ocular nasal discharge} \\
\quad + 0.464 \times \text{nervous signs} \\
\quad + 0.510 \times \text{recumbent posture}.
\]

**Discussion**

In the present study, it was appeared that 145 animal (89.0%) were correctly classified by using first LDA. The average rate of correct classification for *Anaplasma* and *Babesia* was 78.8 and 100%, respectively; while, the second LDA showed that 87.1% of the animals were correctly classified as survivors and non-survivors. The observed clinical signs were commonly

---

**Table 1. Results of two LDA models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Wilks’ Lambda</th>
<th>Chi-square</th>
<th>Degree of freedom</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>First LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Function 1</td>
<td>0.058</td>
<td>309.361</td>
<td>12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Function 2</td>
<td>0.376</td>
<td>105.988</td>
<td>5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Second LDA</td>
<td>0.655</td>
<td>46.744</td>
<td>3</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**Table 2. Loadings (standardized coefficients) and centroids in the discriminant function in two LDA models. Bold: important predictors for the discrimination**

<table>
<thead>
<tr>
<th>Clinical finding</th>
<th>(Predictor)</th>
<th>Loadings (Coefficients)</th>
<th>Loadings (Coefficients)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>First LDA</td>
<td>Function 1</td>
</tr>
<tr>
<td>Oculonasal discharge</td>
<td>1.735</td>
<td>-0.252</td>
<td>0.558</td>
</tr>
<tr>
<td>Hemoglobinuria</td>
<td>0.576</td>
<td>0.295</td>
<td></td>
</tr>
<tr>
<td>Respiratory rate</td>
<td>-0.282</td>
<td>0.728</td>
<td></td>
</tr>
<tr>
<td>Mucous Membrane color</td>
<td>-0.434</td>
<td>0.620</td>
<td></td>
</tr>
<tr>
<td>Lymph node size</td>
<td>-0.495</td>
<td>-0.373</td>
<td></td>
</tr>
<tr>
<td>Bloody feces</td>
<td>-0.894</td>
<td>-0.850</td>
<td></td>
</tr>
<tr>
<td>Nervous signs</td>
<td>0.464</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posture</td>
<td>0.510</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Centroid</th>
<th>First LDA</th>
<th>Function 1</th>
<th>Function 2</th>
<th>Second LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaplasma</td>
<td>6.550</td>
<td>1.490</td>
<td>Survived</td>
<td>2.201</td>
</tr>
<tr>
<td>Babesia</td>
<td>0.663</td>
<td>-1.956</td>
<td>Non survived</td>
<td>-0.235</td>
</tr>
<tr>
<td>Negative to TBDs</td>
<td>-1.342</td>
<td>0.688</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
present with most subjects and the clinical examination of affected animals alone was not successfully revealed the underlying pathogen species. Even a combination of clinical signs and microscopic investigation did not substantially improve the diagnosis, because the clinical signs and symptoms were not identical and greatly confused with many other diseases. Among the tested clinical variables for the first model, the first discriminant functions identified oculonasal discharge, bloody feces and hemoglobinuria as the important predictors for the first function while bloody feces and respiratory rate were the best predictors for the second one for identifying the type of infection, respectively; while nervous cattle with abnormal posture and oculonasal discharge likely to die. Future studies are needed for further confirmation of this predictor for discrimination and classification.

### Table 3. Classification table for the two LDA models

<table>
<thead>
<tr>
<th></th>
<th>First LDA Prediction</th>
<th></th>
<th>Second LDA Prediction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anaplasma</td>
<td>Babesia</td>
<td>Negative to TBDs</td>
<td>Total</td>
</tr>
<tr>
<td>Original Count</td>
<td>26</td>
<td>2</td>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>Babesia</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Negative to TBDs</td>
<td>11</td>
<td>0</td>
<td>108</td>
<td>119</td>
</tr>
<tr>
<td>%</td>
<td>78.8</td>
<td>6.1</td>
<td>15.2</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>9.3</td>
<td>0</td>
<td>90.7</td>
<td>100.0</td>
</tr>
</tbody>
</table>

#: classification table of leave-one-out cross-validation

The results herein demonstrated that discriminant analysis could provide information to increase the understanding of the discrimination among species of TBDs and could help predict a group membership for selection process as well as determining the best predictor for discrimination function. In this context, the first discriminant functions identified oculonasal discharge, bloody feces and hemoglobinuria as important predictors while bloody feces and respiratory rate were the strong predictors in the second function for discrimination of babesiosis and anaplasmosis, respectively; while nervous cattle with abnormal posture and oculonasal discharge likely to die. Future studies are needed for further confirmation of this predictor for discrimination and classification.

### Conclusion

The results herein demonstrated that discriminant analysis could provide information to increase the understanding of the discrimination among species of TBDs and could help predict a group membership for selection process as well as determining the best predictor for discrimination function. In this context, the first discriminant functions identified oculonasal discharge, bloody feces and hemoglobinuria as important predictors while bloody feces and respiratory rate were the strong predictors in the second function for discrimination of babesiosis and anaplasmosis, respectively; while nervous cattle with abnormal posture and oculonasal discharge likely to die. Future studies are needed for further confirmation of this predictor for discrimination and classification.
Acknowledgment

The authors would like to thank Dr. Mohamed El-Beskawy for his kind support in sample collection. The authors appreciate the great help of Professor Dr. Heinrich Neubauer, the director of the Friedrich Loeffler Institute (FLI), Federal Research Institute for Animal Health, Jena, Germany and Dr. Helmut Hotzel and Dr. Herbert Tomaso for their kind help and support for characterizing the samples.

References


13) McClelland G, Melendy J. Use of parasites as tags in delineating stocks of Altaic cod (Gadusmorhua) from the southern Gulf of St. Lawrence and the Cape Breton Shelf. Fisheries Research 107, 233–238, 2011.


